1. Data Cleaning/pre-processing

**Step 1:** Clean the data

**Step 2:** Find the event season

**Step 3:** Find the difference between the start date and the purchase date (days and weeks)

**Step 4:** Find a general correlation to know which columns to drop (if necessary)

2. EDA graphs and table (I chose to deal with seasons because event names are over 100, and event types are incomplete)

**Step 5:** Find the total ticket purchased per week and patterns for events in each season (to check if there is a pattern of purchase for events in each season)

**Step 6:** A table showing the number of weeks it took to sell out tickets for each event in each season (days between the first and last purchase)

**Step 7:** A table showing how close the “sell-out” weeks are to the event date (days between the last purchase and the event start date)

**Step 8:** A table showing the event date, first ticket purchase date, with the total number of tickets eventually sold (to show how long before the event they started selling the ticket)

3. Hypothesis

**Step 9:** Find the correlation between the start date and purchase date

**Step 10:** Find the correlation between the season and the total tickets purchased

**Step 11:** Find the correlation between the season and the number of weeks it took to sell out the tickets (does it happen by chance or by event or by season?)

**Step 12:** Relationship between event date and first ticket purchase date, with the total number of tickets eventually sold (this is to advise on how quickly they should start advertising for the event)

**Step 13:** Cancellation based on seasons (to check if people shift their interest based on season)

4. Prediction

**Step 14:** Predict future weekly purchases with ARIMA or any other time series

**Step 15:** Predict annual purchases with a Facebook prophet model

**Step 16:** Based on the total number of weeks it took to sell out tickets, predict future events in the same season

**Step 17:** Based on the total number of weeks remaining to the event after selling out the ticket, predict future events in the same season

5. Recommendations

* If we have the target/expected audience, we can predict further
* Knowing where each user clicked the purchase link will also help predict further.
* If we have the total number of people that finally attended, we can find a correlation between purchased and attended.
* If we have a complete event type, we may be able to predict the behavior per event type in each season
* If we have when the event advert started, we may be able to predict further

Analysis/Algorithms:

1. Data cleaning types
2. Season from excel or python pandas
3. Date diff in excel or python pandas
4. Grouping data into seasons
5. Plotting time series graphs
6. Algorithms for finding correlations
7. Algorithms for predictions

Proposed document format

* Introduction
* Methodologies (data used, functions used, algorithms used, and WHY???)
* Findings/Solution (EDA Graphs, prediction Graphs, and tables)
* Conclusions
* Recommendation
* Appendix